

Arima as a forecasting tool for water quality time series measured with UV-Vis spectrometers in a constructed wetland

• Nathalie Hernández • Julio Camargo • Fredy Moreno •
Pontificia Universidad Javeriana, Colombia

• Leonardo Plazas-Nossa* •
Pontificia Universidad Javeriana, Colombia | Universidad Distrital Francisco José de Caldas, Colombia

*Corresponding autor

• Andrés Torres •
Pontificia Universidad Javeriana, Colombia

DOI: 10.24850/j-tyca-2017-05-09

Abstract

Hernández, N., Camargo, J., Moreno, F., Plazas-Nossa, L., & Torres, A. (September-October, 2017). Arima as a tool to predict water quality using time series recorded with UV-Vis spectrometers in a constructed wetland. *Water Technology and Sciences* (in Spanish), 8(5), 127-139.

The prediction of water quality plays a crucial role in discussions about urban drainage systems, given that the integrated management of this resource is required in order to meet human needs. The present paper uses Arima (Autoregressive Integrated Moving Average) to predict influent and effluent water quality in a constructed wetland, as well as its pollutant removal efficiency. The wetland is located on the campus of the Pontificia Universidad Javeriana in Bogotá, Colombia. Arima prediction values were based on time series obtained with UV-Vis spectrometry probes. These predictions were found to be adequate for the first 12 hours of the water quality time series for the three data sets analyzed: influent, effluent, and efficiency. Overall, none of the data had prediction errors over 15%. In separate analyses of the relative predictive errors in influent and effluent values, they were found to be less significant for UV wavelengths than for the visible range (Vis). In addition, the variability in this type of error was less for the UV range than for the Vis range, which indicates that Arima is a suitable prediction method for analyzing pollutants that fall in the UV range.

Keywords: Forecasting methods, time series analysis, UV-Vis spectrometry, water quality, wetland.

Resumen

Hernández, N., Camargo, J., Moreno, F., Plazas-Nossa, L., & Torres, A. (septiembre-octubre, 2017). Arima como herramienta de pronóstico para la calidad del agua con series de tiempo registradas con espectrómetros UV-Vis en un humedal construido. *Tecnología y Ciencias del Agua*, 8(5), 127-139.

Cuando se discute sobre el tema de los sistemas de drenaje urbano, el pronóstico de la calidad del agua juega un papel crucial, dado que la gestión integrada de este recurso es necesaria para satisfacer las necesidades humanas. El presente artículo aplica Arima (Autoregressive Integrated Moving Average) para pronosticar la calidad del agua afluente y efluente, además de la eficiencia de eliminación de contaminantes, en un humedal construido ubicado en el campus de la Pontificia Universidad Javeriana en Bogotá, Colombia. Los valores del pronóstico de Arima se basan en series temporales obtenidas por sondas de espectrometría UV-Vis. Los pronósticos basados en Arima son adecuados para las primeras 12 horas de la serie de tiempo de calidad del agua, y para las tres series de tiempo analizadas: afluente, efluente y eficiencia. En general, los errores de pronóstico no sobrepasaron el 15% para ninguno de los datos observados. Análisis separados del afluente y del efluente respecto a los errores de pronóstico relativos resultantes prueban ser menos significativos para las longitudes de onda UV que para el rango visible (Vis). Asimismo, para el rango UV, este tipo de error presenta una menor variabilidad que la de la gama Vis, un resultado que sugiere que Arima es un método de pronóstico adecuado cuando se analizan contaminantes que caen en el rango UV.

Palabras clave: análisis de series de tiempo, calidad de aguas, espectrometría UV-Vis, humedal, métodos de pronóstico.

Received: 03/06/2015
Accepted: 04/04/2017

Introduction

Water quality monitoring has become an indispensable part of the management of urban drainage systems given that climate variables or contaminant loads can quickly alter water quality. Normally carried out via sampling, quality control for these systems entails the collection, transportation and laboratory analysis of field samples. More often than not, these laboratories are not found in the same place as the sample collection site. Here, the spatiotemporal representation achieved by sampling must be mentioned in tandem with the problems it presents, such as the systematic errors produced by the laboratory equipment (Plazas-Nossa & Torres, 2014).

In response to this issue, time and money have been invested in online sensors for water quality monitoring; these sensors offer the possibility of real-time measurements (Qin, Gao, & Chen, 2011; Zamora & Torres, 2014). Optic and electronic development has brought with it advances in Ultraviolet (UV) and Visible (Vis) spectrometry (UV-Vis), a field focused on producing small-scale robust sensors that register light attenuation (absorbance) and provide continuous water quality results (at a rate of one signal per minute) (Plazas-Nossa & Torres, 2014). One of the primary advantages of this type of sensor is its ability to simultaneously track various parameters with a single measuring device (Gruber, Bertrand-Krajewski, De Bénédittis, Hochedlinger, & Lettl, 2006; De Sanctis, Del Moro, Levantesi, Luprano, & Di Iaconi, 2016; Vanacker, Wezel, Arthaud, Guérin, & Robin, 2016). UV-Vis spectrometry has proven to be useful for water quality measuring, particularly in wastewater treatment plants, where it is used at different treatment stages to evaluate both contaminant loads and removal efficiency of organic material, nitrates, nitrites and Total Suspended Solids (TSS) (Plazas-Nossa & Torres, 2014).

Water quality predictions for urban sanitation hydro-systems take on added significance when attempting to forecast the future behavior

of different contamination determinants to grant decision-makers the tools with which to follow the appropriate preventive or corrective steps related to water quality management. The pertinent scientific literature reports experiences with water quality prediction (Faruk, 2010; Yan, Zou, & Wang, 2010; Halliday *et al.*, 2012; Campisano, Cabot, Muschalla, Pleau, & Vanrolleghem, 2013; García *et al.*, 2015; Garcia *et al.*, 2016; Hornsby, Ripa, Vassillo, & Ulgiati, 2016). There are also cases of classic prediction models employed in a wide array of water quality studies, such as Autoregressive Moving Average (ARMA) or the Box-Jenkins Autoregressive Integrated Moving Average (Arima) (Lehemann & Rode, 2001; Faruk, 2010; Abaurrea, Asín, Cebrián, & García-Vera, 2011; Widowati, Purnomo, Koshio, & Oktaferdian, 2016; Brentan, Luvizotto, Herrera, Izquierdo & Pérez-García, 2017).

However, the literature provides no evidence of the application of these methods for predicting UV-Vis spectrometry time series with short acquisition phases (on the order of one spectrum per minute); moreover, few cases speak on the subject from the point of view of other methods, such as the Discrete Fourier Transform (DFT) (Plazas-Nossa & Torres, 2013) or Artificial Neural Networks (ANN) (Plazas-Nossa, Avila, & Torres, 2017a).

Water quality prediction also facilitates the recycling of rain water, especially regarding supporting the decision-making process related to the allocation of funds towards the development of rainwater harvesting infrastructure.

The constructed wetland under study for the present paper includes a continuous water quality testing system that looks at affluent and effluent with UV-Vis spectrometers (Spectro::lyser™) (Galarza-Molina, Torres, Moura, & Lara-Borrero, 2013). Observed water quality presents temporal fluctuations, a situation that can be attributed to the presence of substances such as inorganic ions, heavy metals and pathogenic microorganisms. With an eye towards creating a system with real-time control that maximizes the amount of recycled rainwater, a predictive

tool has been proposed to ensure the quality of affluent and effluent water. On the whole, *in situ* time series display stochastic behavior. The manipulation and analysis of this data obviously becomes complex, leading the authors of the present article to rely on Arima, a method that models time series with or without component tendencies or seasonal variations. Also, Arima lends itself to forecasting (Shyh-Jier & Kuang-Rong, 2003; Faruk, 2010; Abaurrea et al., 2011; Widowati et al., 2016; Brentan et al., 2017).

Materials and methods

The Pontificia Universidad Javeriana in Bogotá, Colombia serves as the site of this case study. The campus's constructed wetland (regulator tank) is four meters wide and 20.3 meters long. On its surface is a vegetative layer of papyrus with the following characteristics: the first zone is 4 m by 6.72 m, in one inch of gravel; the second is 4 m by 6.97 m in 0.75 inch of gravel; the third and final is 4 m by 6.62 m in 0.5 inch of gravel. The wetland is designed for subsurface flow to take in overflow from the Guillermo Castro Building (a parking lot) and the Néstor Santacoloma Building (the building that houses Oncology) (Galarza-Molina et al., 2013).

Spectro::lyser™ UV-Vis waterproof spectrometers, 65 centimeters long by 44 millimeters wide, are used to conduct the present research. Primarily used to record light attenuation (absorbance) on-line continuously (one signal per minute), these spectrometers are equipped with a xenon light of wavelength 200 nm to 750 nm at 2.5 nm intervals (Plazas-Nossa & Torres, 2013; Plazas-Nossa, Torres, Gruber, & Hofer, 2014). The spectrometers are located at the input (affluent) and output (effluent) of the constructed wetland. Table 1 details the wavelengths (200 to 745 nm) for which the contamination determinants consider in the present study are relevant (e.g. nitrates, nitrites, chemical oxygen demand and biochemical oxygen demand) (Plazas-Nossa et al., 2014).

Drawn from the wetland's affluent and effluent, the data presents in the table 1 is taken continuously from 12:00 a.m. on March 6th, 2014 to 6:10 a.m. on March 21st, 2014 at one-minute intervals. In total, 21251 pieces of data are recorded. Spectrometers registered 219 total wavelengths for affluent and 214 for effluent. The difference between these two wavelengths (219 and 214) can be explained by the characteristics of the sensors themselves given by the sensor's manufacture parameters. Due to

Table 1. Wavelength and contamination determinants (Source: Plazas-Nossa et al., 2014).

Spectrum	Parameters	Wavelengths Ranges (nm)
UV	NO ₂ Nitrites and NO ₃ Nitrates, Detergents (benzene forms) at 225 nm	200-250
	COD-1 Acetone 266 nm	252.5-267.5
	Phenols Acetaldehyde 277 nm	270-286
	COD-2 (Phenols), presence of hypochlorite ion 290 nm	287.5-357.5
	Formaldehyde	360-380
VISIBLE	DOC	382.5-427.5
	Violet	430-477.5
	Blue	480-537.5
	Green	540-577.5
	Yellow	580-617.5
	Orange	620-647.5
	Red	650-687.5
	TSS	690-745

interference of this nature, the present study only accounts for the first 214 wavelengths which are captured by either the input or output sensors (200 nm to 735 nm). To forestall the problem of missing or atypical data, raw data is filtered with a combination of Winsorizing (Ko & Lee, 1991; Liu, Shah, & Jiang, 2004) and DFT (Plazas-Nossa, Torres, Gruber, & Hofer, 2017b). Winsorizing constitutes a technique specifically designed to handle such cases of data filtering.

Principal Component Analysis (PCA) (Juhos, Makra, & Tóth, 2008; Shlens, 2009; Krawczak & Szkatula, 2014) separates the 214 wavelengths for the 21251 pieces of absorbance data into principal components, gathering the most information possible (more than 95%) such that they lessen computational strain. What's more, PCA makes data series reconstruction possible; in other words, armed with principal components, it is possible to reconstruct the steps culminating in the 214 wavelengths. The authors of this study added Arima (Box, Jenkins, & Reinsel, 1993) to the use of PCA to remove data trends and variations. Thus, this combination achieves stationarity. PCA and Arima form the base from which different forecasting times are generated. Wetland retention time, a key aspect when determining retention efficiency, is determined *via* the Cross-correlation Function (CCF), which is applied to the first principal components of affluent and effluent. This analysis provides the constructed wetland/regulator tank's retention time, a factor taken into account for the effluent data study. The two data sets (affluent and effluent) are split in 2/3 for calibration and the rest for validation. While the calibration data consists of applying Arima to the three principal PCA components (i.e. the work done by this study), the validation data is comprised of the observed data, a necessary step for checking the "validity" of the forecast time series data. Instead of directing attention solely on affluent and effluent, data analysis is also done on wetland efficiency (the ability of the wetland to remove contamination determinants). Therefore, data analysis can be broken down into three phases: (i) PCA and Arima

to pre-evaluate the information and generate predictions for the principal components; (ii) PCA and Arima for the selected principal components are followed by reverse PCA to reconstruct the prediction for all 214 wavelengths based on Arima (of the principal components); (iii) PCA, Arima and reverse PCA carry out on the input/output data efficiency to forecast in terms of wetland efficiency.

Prediction results for the three analytical steps, and two groups (input/output), produce an average of forecasting of confidence intervals between 80% and 95%, it is a range in which each type of data analysis falls (time behaviour of principal components, of each wavelength and of input/output data efficiency). The data is then analyzed using a sort of control: forecast data is checked (calibration) against the real data (validation data) for each type of analysis. In turn, a global analysis of the series is facilitated, as well as a point-by-point comparison of forecast and real values and a calculation of relative and absolute errors for each value. Seeing as the data distribution is unknown, dispersion and box plot are determined. This last step sums up the total behavior for wavelength or principal component for each time step.

Results obtained allow a final analysis for each wavelength range (contamination determinants displayed in table 1) to verify if the observed data matches—or not—the Arima prediction. Also, acceptable forecasting time and wetland efficiency are determined.

Data analysis uses the mathematical software R (R Development Core Team 2014) and the Forecast Package: Forecasting Functions for Time Series and Linear Models (Hyndmann & Khandakar, 2008) for Arima and CCF.

Results and discussion

For the input and output time series, 21251 pieces of total data are obtained. Of these 21251, the first 14026 go towards calibration, with the rest for validation (7225)—the 2/3:1/3 proportion previously mentioned. The time limit for the maximum forecast is 7225 minutes (120 hours),

a direct match to the amount of validation data. CCF concludes that wetland retention time is 19 minutes, with the highest correlation found to be 0.22.

For initial data analysis, real and forecast time series for the first three principal components were compared. Arima is the used method for forecasting the affluent and effluent time series. As far as relative errors are concerned, the former contains relative errors with a trend towards growth, reflecting increases in the prediction forecasting time (except for minutes 717 minutes and 737, at which spikes are observed). For the latter, relative errors correlate with the mean forecast, jumping when the forecasting time hits 3 500 minutes.

Results stemming from the second type of analysis are as follows: affluent data confirms that Arima prediction trends and the validation series exhibit relative errors (figure 1); both sets trend towards a growth approaching 80%, a finding best understood as a difference in precision between the forecast and the real value. The highest percentage of relative errors occurs

between 500 nm and 735 nm, which correspond to the visible spectrum (TSS and turbidity). Thus, Arima forecasts best fit the parts the UV part of the spectrum (organic material-related contaminants), as illustrated by figure 1. For wavelengths 205, 207.5, 215, 217.5, 220, 225, relative error peaks at 14%, whereas absolute errors display their highest values—6.6 absorbance units (AU)—at wavelengths 205, 207.5, 215, 217.5, 220, 225. Here, outliers are present. Absolute errors for the remaining wavelengths turn out to be insignificant.

Analyses performed on Arima trend predictions and validation series about the forecasting time reveals relative errors from 0 to 80%, peaking around 40% between 719 and 739 minutes (figure 2). Absolute errors mirror this behavior, though their maximum value (in the neighborhood of 2.5 AU) appears not in the final minutes but between minutes 719 and 739.

Having presented results for affluent data, attention now shifts towards effluent data, i.e. what leaves the system. Figure 3 portrays Arima prediction trends and validation series obtained

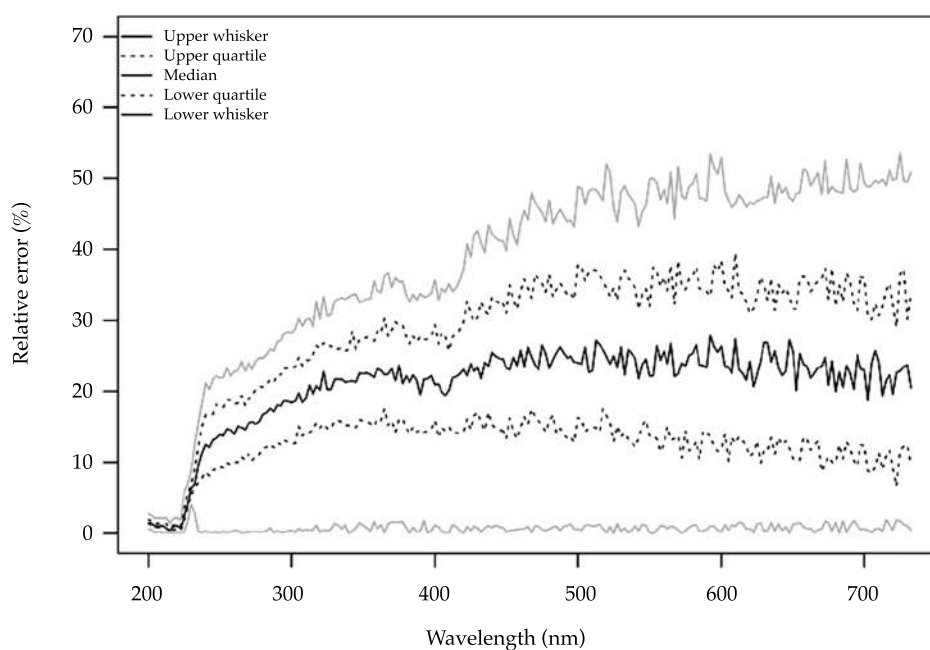


Figure 1. Wavelength and relative error (affluent/input data).

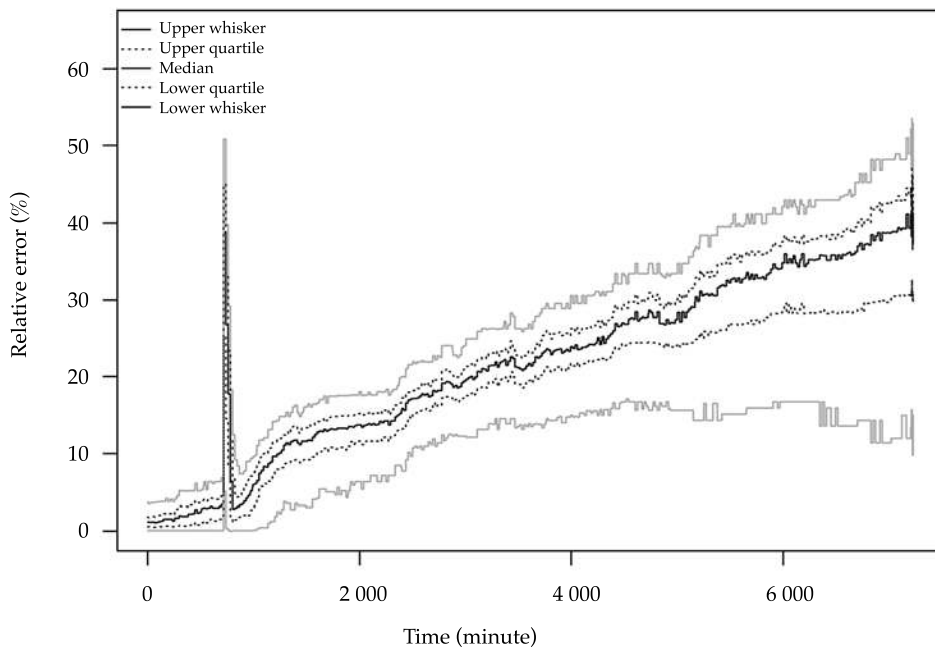


Figure 2. Forecasting time and relative error (affluent/input data).

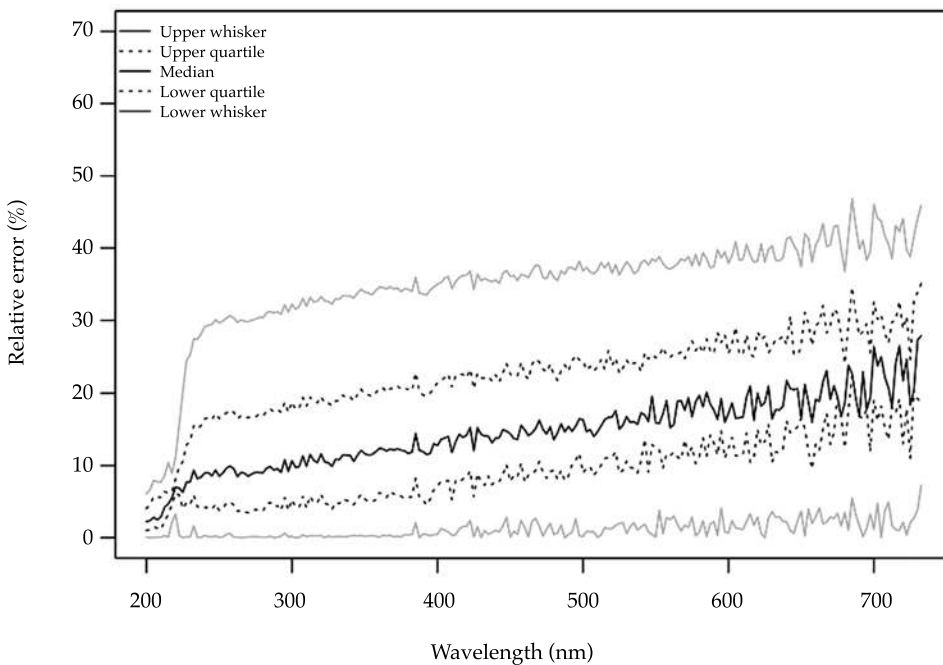


Figure 3. Wavelength and relative error (effluent/output data).

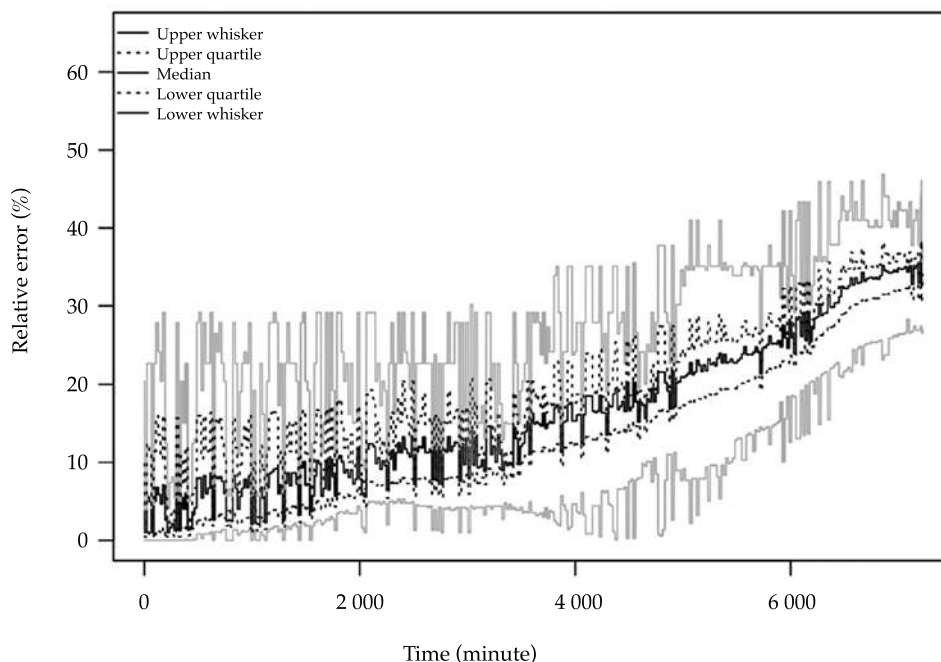


Figure 4. Forecasting time and relative error (effluent/output data).

for wavelengths. For this data, relative errors display a growth trend, closing in values of 50% at higher wavelength spectra (roughly 680 nm to 735 nm). Similarly, these errors steadily increase, correlating an increase in wavelength, from 220 nm to 735 nm. Absolute error is in the range of 12.6 AU, with its lowest value displayed in the higher wavelengths, a finding which confirms Arima's forecasting benefits in terms of wavelength predictions related to organic contamination (graph not shown in this paper).

Arima prediction trend and validation time series relating to the time limit see relative and absolute errors increase from 0% to 50% for relative errors and 0 AU to 12 AU for absolute errors, respectively. Variability levels observed for relative errors display is especially high (figure 4).

For the third, and final, stage of data analysis, shown in figure 5, the wetland's contaminant removal efficiency is compared to wavelengths. Figure 5 evinces the negative efficiency (maximum value of -1) for the 225 nm to 435 nm

range when it comes to eliminating nitrates and nitrites, acetone, phenols, hypochlorite, formaldehyde, COD, total organic carbon, benzenes and toluene (readers are referred to table 1). This indicates that effluent absorbance values double (in magnitude) affluent absorbance values. As a result, it is plausible to conclude that the wetland does not efficiently remove contaminants corresponding to these wavelengths. This inefficiency might be argued to be rooted in the fact that papyrus (in charge of contaminant removal) were not originally planted in the wetland. Rather, they were planted and raised elsewhere and then replanted in the wetland. Despite being a seemingly innocuous observation, this difference plays a large role in the papyrus' ability to retain contaminants (nitrates and nitrites), considering that the highest retention capacity manifests during growth and development processes. Furthermore, the wetland has not been trimmed/pruned, causing some plants that have already completed their life cycle to decompose and subsequently release

their nutrients and/or retained contaminants into the wetland.

Not surprisingly, then, from wavelength 437.5 nm on, the wetland's efficiency improves, effectively retaining contaminants associated with these wavelengths (those of the visible part of the UV-Vis spectrum), such as turbidity, color and TSS at rates of up to 70% (see figure 5). In essence, the wetlands removal of contamination determinants thrives for turbidity and TSS.

In figure 6, a trend of temporal weakening of the efficiency is observed, a situation attributable to the contaminant saturation of the vegetative layer. Eight peaks can be seen throughout the prediction limit, with peaks taking place every 29 hours (approximately). While the cause of this behavior has not been sufficiently explored, it could be the product of papyrus life-cycles and/or the result of contaminants during rain events.

Figure 7 lays out the fluctuations of relative errors for each wavelength versus the average

of the forecast (predicted) and observed (validated) series. Between wavelengths 200 nm and 210 nm, the average forecast is plagued by high relative error; these errors are as high as 2 000%. Almost as striking is the fact that from wavelength 455 nm on, we observe errors around 1 000%.

Nonetheless, observed absolute errors are quite low (less than 1 AU) for wavelengths with high levels of relative errors (see figure 8). Therefore, despite increased relative errors (figure 7), the proposed forecasting tool provides accurate results for the entire UV-Vis range.

Relative and absolute error analysis for the average of prediction and control series takes the prediction limit into account (figures 9 and 10). These two graphs visually represent the error variability for distant (i.e. farther ahead in terms of time) predictions. A peak pops up at minute 717, suggesting that the forecast is acceptable before said time.

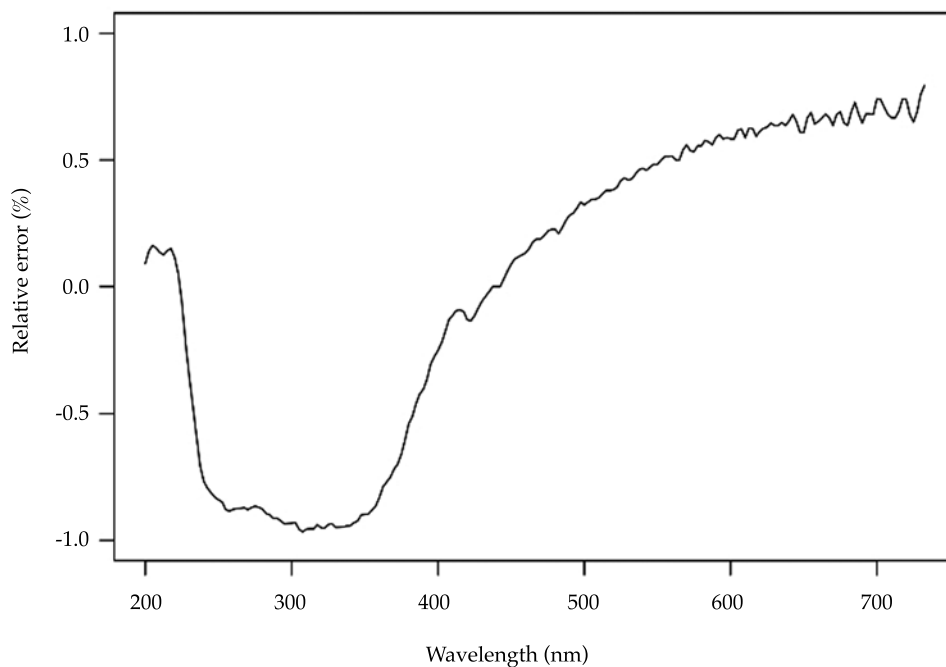


Figure 5. Wavelength and median absorbance efficiency.

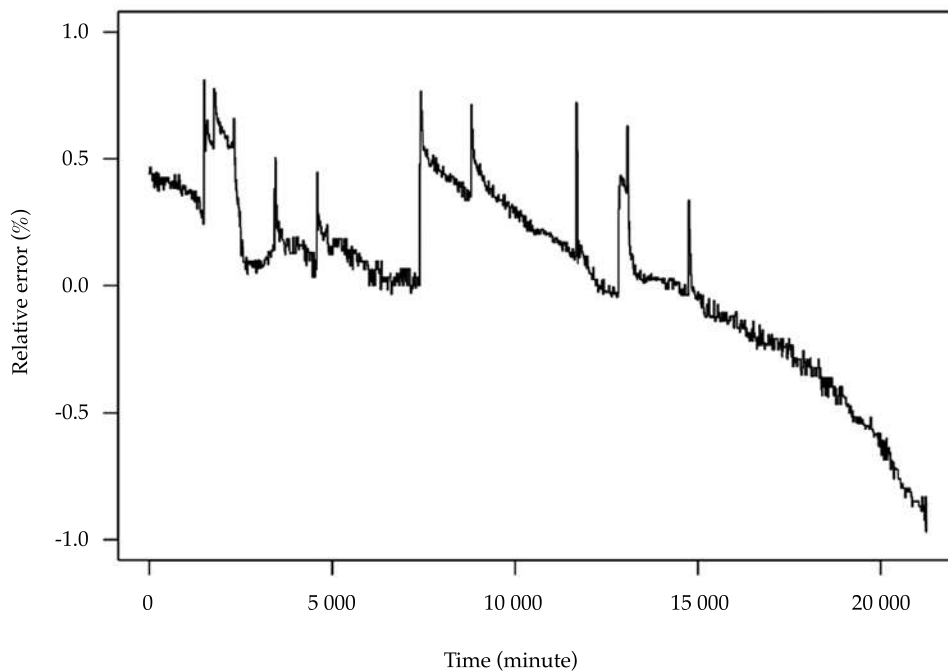


Figure 6. Median of absorbance efficiency and total time range.

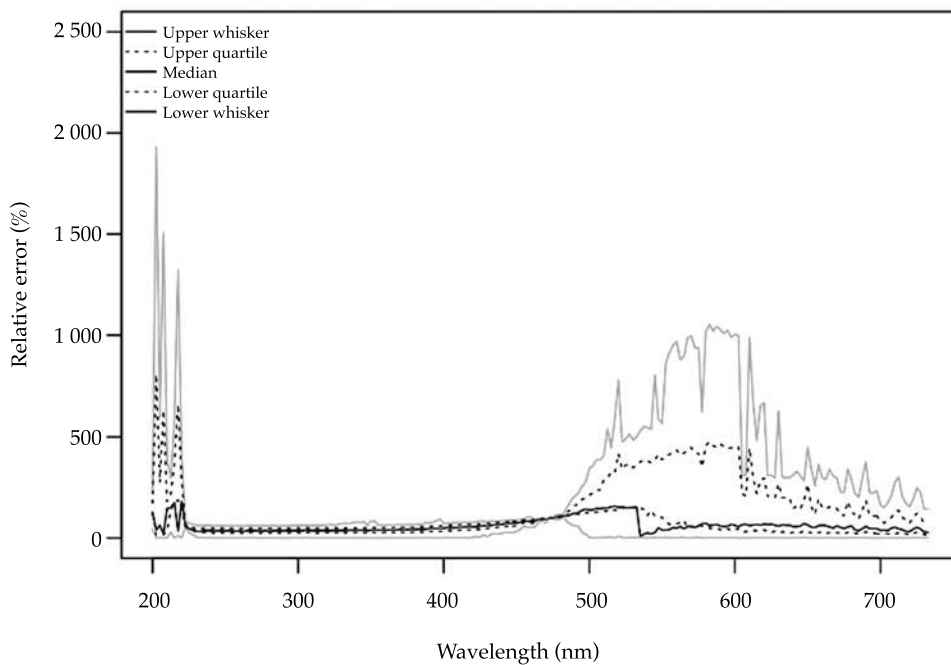


Figure 7. Wavelength and relative error (efficiency).

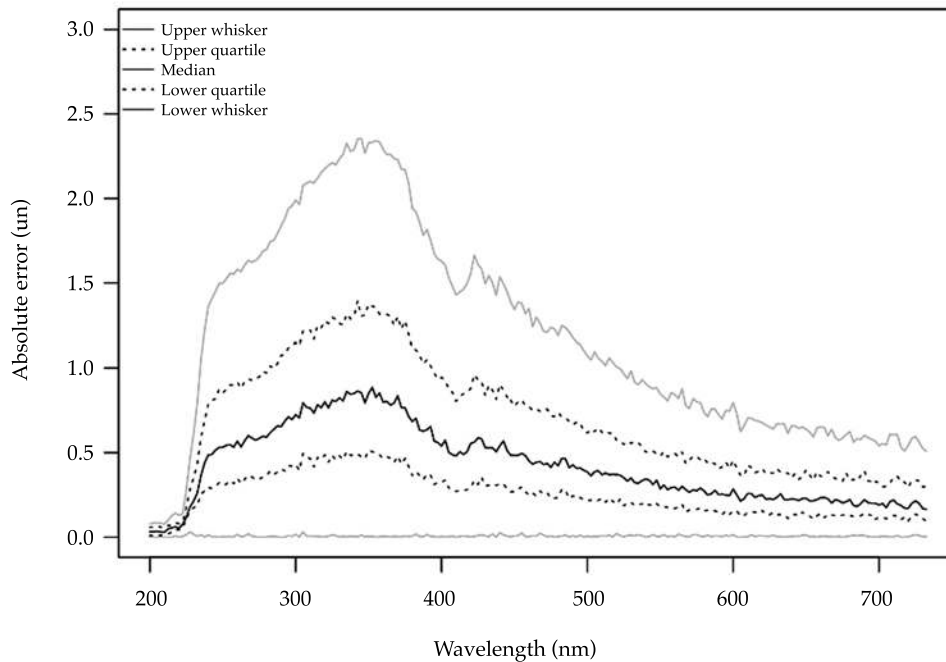


Figure 8. Wavelength and absolute error (efficiency).

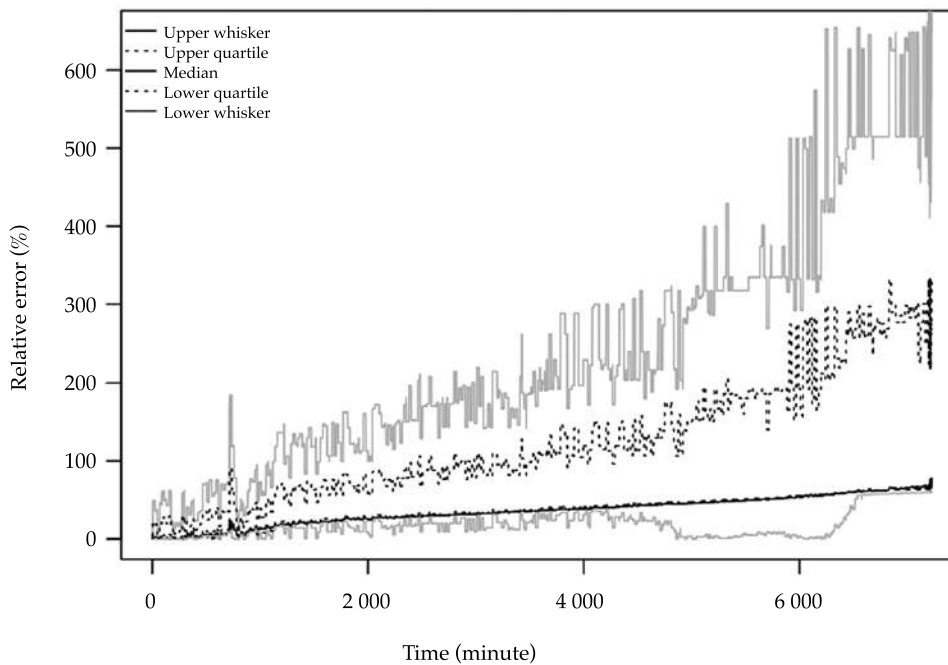


Figure 9. Relative error and forecasting time (efficiency).

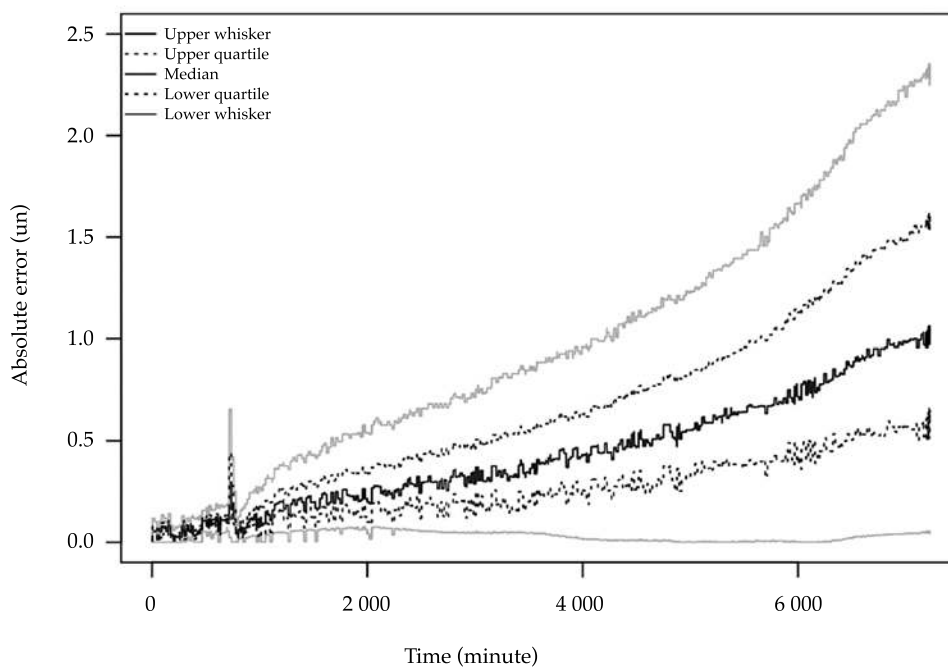


Figure 10. Absolute error and forecasting time (efficiency).

Conclusions

In this paper, the Arima methodology is used to forecast water quality time series measured with UV-Vis sensors in a constructed wetland located on the campus of the Pontificia Universidad Javeriana. Measurements for the affluent (input), effluent (output) and efficiency of this constructed wetland are analyzed.

Arima-based predictions appropriately forecast the first 12 hours of the water quality time series for the three data sets analyzed: affluent, effluent and efficiency. Prediction errors did not exceed 15% for any of the observed data. The accuracy of said predictions is based on a comparison to a control (validation) series arrived at using field-observed data.

Separate analyses of affluent and effluent testify to the fact that relative prediction errors resulting from Arima prove to be less significant for UV wavelengths than for the visibility (Vis) range. This refers to wetland's improved capacity for handling turbidity and TSS versus

other contaminants, such as nitrates, nitrites and toluene. Likewise, for the UV range, these errors exhibit less variability than for the Vis range. Naturally, such an outcome suggests that Arima is a valuable prediction method when discussing contaminants that fall in the UV range.

References

- Abaurrea, J., Asín, J., Cebrián, A., & García-Vera, M. (2011). Trend analysis of water quality series based on regression models with correlated errors. *Journal of Hydrology*, 400, 341-352.
- Box, G., Jenkins, G., & Reinsel, G. (1993). *Time series analysis: Forecasting and control* (3rd ed.). Englewood Cliffs: Prentice Hall.
- Brentan, B., Luvizotto, E., Herrera, M., Izquierdo, J., & Pérez-García, R. (2017). Hybrid regression model for near real-time urban water demand forecasting. *Journal of Computational and Applied Mathematics*, 309, 532-541.
- Campisano, A., Cabot, J., Muschalla, D., Pleau, M., & Vanrolleghem, P. (2013). Potential and limitations of modern equipment for real time control of urban wastewater systems. *Urban Water Journal*, 10(5), 300-311.

- De Sanctis, M., Del Moro, G., Levantesi, C., Luprano, M., & Di Iaconi, C. (2016). Integration of an innovative biological treatment with physical or chemical disinfection for wastewater reuse. *Science of the Total Environment*, 543, 206-213.
- Faruk, D. (2010). A hybrid neural network and Arima model for water quality time series prediction. *Applications of Artificial Intelligence*, 23, 586-594.
- Galarza-Molina, S., Torres, A., Moura, P., & Lara-Borrero, J. (2013). CRIDE: Multi-criteria analysis tool for decision making support for rainwater harvesting in the university campus of Pontificia Universidad Javeriana. *Proceedings of 5th IWA International Conference on Benchmarking and Performance Assessment of Water Services*. Medellín, Colombia, 9-12 April.
- García, L., Barreiro-Gomez, J., Escobar, E., Téllez, D., Quijano, N., & Ocampo-Martinez, C. (2015). Modeling and real-time control of urban drainage systems: A review. *Advances in Water Resources*, 85, 120-132.
- García, X., Barceló, D., Comas, J., Corominas, L., Hadjimichael, A., Page, T., & Anuña, V. (2016). Placing ecosystem services at the heart of urban water systems management. *Science of the Total Environment*, 563-564, 1078-1085.
- Gruber, G., Bertrand-Krajewski, J.-L., De Bénédittis, J., Hochedlinger, M., & Lettl, W. (2006). Practical aspects, experiences and strategies by using UV/Vis sensors for long-term sewer monitoring. *IWA Water Practice & Technology*, 1(1), 1-8.
- Halliday, S., Wade, A., Skeffington, R., Neal, C., Reynolds, B., Rowland, P., Neal, M., & Norris, D. (2012). An analysis of long-term trends, seasonality and short-term dynamics in water quality data from Plynlimon, Wales. *Science of the Total Environment*, 434, 186-200.
- Hornsby, C., Ripa, M., Vassillo, C., & Ulgiati, S. (2016). A roadmap towards integrated assessment and participatory strategies in support of decision-making processes. The case of urban waste management. *Journal of Cleaner Production*, 142(1), 157-172.
- Hyndmann, R., & Khandakar, Y. (2008). Automatic time series forecasting: The forecast package for R. *Journal of Statistical Software*, 27(3), 1-22.
- Juhos, I., Makra, L., & Tóth, B. (2008). Forecasting of traffic origin NO and NO₂ concentrations by Support Vector Machines and neural networks using Principal Component Analysis. *Simulation Modelling Practice and Theory*, 16, 1488-1502.
- Ko, S.-J., & Lee, Y. (1991). Theoretical analysis of winsorizing smoothers and their applications to image processing (pp. 3001-3004). *Proceedings of ICASP-1991 International Conference on Acoustics, Speech, and Signal Processing*, 4, 14-17 April 1991.
- Krawczak, M., & Szkatula, G. (2014). An approach to dimensionality reduction in time series. *Information Sciences*, 260, 15-36.
- Lehemann, A., & Rode, M. (2001). Long-term behavior and cross-correlation water quality analysis of the river Elbe, Germany. *Water Research*, 35(9), 2153-2160.
- Liu, H., Shah, S., & Jiang, W. (2004). Online outlier detection and data cleaning. *Computers and Chemical Engineering*, 28(9), 1635-1647.
- Plazas-Nossa, L., & Torres, A. (2013). Fourier analysis as a forecasting tool for absorbance time series received by UV-Vis probes installed on urban sewer systems. *Proceedings of 8th International Conference on Planning & Technologies for Sustainable Urban Water Management Novatech-2013*, 23-27 June 2013, Lyon-France.
- Plazas-Nossa, L., & Torres, A. (2014). Comparison of Discrete Fourier Transform (DFT) and principals components analysis/DFT as forecasting tools of absorbances time series received by UV-Visible probes installed in urban sewer systems. *IWA Water Science & Technology*, 69(5), 1101-1107.
- Plazas-Nossa, L., Torres A., Gruber, G., & Hofer, T. (2014). PCA/ANN for forecasting of absorbance time series received by UV-Vis probes installed in urban sewer systems. *Proceedings of 13th IAHR/IAW International Conference on Urban Drainage, ICUD-2014*. 7-12 September 2014, Kuching-Malaysia.
- Plazas-Nossa, L., Ávila, M., & Torres, A. (2017a). Detection of outliers and imputing of missing values for water quality UV-Vis absorbance time series. *Revista Ingeniería*, 22(1), 126-139.
- Plazas-Nossa, L., Torres A., Gruber, G., & Hofer, T. (2017b). Forecasting of UV-Vis absorbance time series using artificial neural networks combined with principal component analysis. *IWA Water Science & Technology*, 75(4), 765-774.
- Qin, X., Gao, F., & Chen, G. (2011). Wastewater quality monitoring system using sensor fusion and machine learning techniques. *Water Research*, 46, 1133-1144.
- Shlens, J. (2009). *A tutorial on principal component analysis*. La Jolla, USA: Salk Institute for Biological Studies.
- Shyh-Jier, H., & Kuang-Rong, S. (2003). Short-term load forecasting via ARMA model identification including non-Gaussian process considerations. *IEEE Transactions on Power Systems*, 18(2), 673-679.
- Vanacker, M., Wezel, A., Arthaud, F., Guérin, M., & Robin, J. (2016). Determination of tipping points for aquatic plants and water quality parameters in fish pond systems: A multi-year approach. *Ecological Indicators*, 64, 39-48.
- Widowati, A., Purnomo, S., Koshio, S., & Oktaferdian, V. (2016). Implementation of Arima model to asses seasonal variability macrobenthic assemblages. *Aquatic Procedia*, 7, 277-284.
- Yan, H., Zou, Z., & Wang, H. (2010). Adaptive neuro fuzzy inference system for classification of water quality status. *Journal of Environmental Sciences*, 22(12), 1891-1896.

Zamora, D., & Torres, A. (2014). Method for outlier detection: a tool to assess the consistency between laboratory data and ultraviolet-visible absorbance spectra in wastewater samples. *Water Science and Technology*, 69(11), 2305-2314.

Author's institutional address

M.I. Nathalie Hernández
Ing. Julio Camargo
Ing. Fredy Moreno
Dr. Andrés Torres

Pontificia Universidad Javeriana
Facultad de Ingeniería
Grupo de Investigación Ciencia e Ingeniería del Agua
y el Ambiente
Carrera 7 # 40-62
Tel.: 3208 320, ext. 5553
Bogotá, COLOMBIA
nathalie_hernandez@javeriana.edu.co
julio.camargo@javeriana.edu.co
fmoreno@acueducto.com.co
andres.torres@javeriana.edu.co

Dr. Leonardo Plazas-Nossa

Universidad Distrital Francisco José de Caldas
Facultad de Ingeniería
Carrera 7 # 40-53
Tel.: 3239 300, ext. 2405
Bogotá, COLOMBIA
lplazasn@udistrital.edu.co
plazas-l@javeriana.edu.co